Research Paper

International Law Programme

September 2022

Refugee protection in the artificial intelligence era

A test case for rights

Madeleine Forster

CHATHAM HOUSE

VISIT…

Chatham House, the Royal Institute of International Affairs, is a world-leading policy institute based in London. Our mission is to help governments and societies build a sustainably secure, prosperous and just world.

Summary

- Government and private sector interest in artificial intelligence (AI) for border security and for use in asylum and immigration systems is growing. Academics and civil society are calling for greater scrutiny of legal, technological and policy developments in this area. However, compared to other high-risk environments for AI, this sector has received little policy attention.
- Whether governments can adopt AI and meet human rights obligations in asylum and immigration contexts is in doubt, particularly as states have specific responsibilities towards persons seeking refugee and humanitarian protection at national borders.
- The risks include potentially significant harm if AI systems lead (or contribute) to asylum seekers being incorrectly returned to their country of origin or an unsafe country where they may suffer persecution or serious human rights abuses – a practice known as 'refoulement'. The use of AI in asylum contexts also raises questions of fairness and due process.
- Some reasons for optimism include recent efforts at responsible innovation. This involves governments focusing their efforts to deploy AI in parts of asylum and related decision-making processes deemed less likely to create tension with domestic and international legal principles.
- However, the restrictive and changeable nature of refugee and immigration policy in many countries today, as well as systemic challenges around fairness and access to rights, creates significant obstacles to human rights-compliant AI. It also creates significant obstacles to community and private sector participation in responsible and collaborative AI development.
- Emerging AI principles and safeguards (e.g. human control, transparency, algorithmic impact assessments) that build on good governance principles will be relevant to future development of systems and policies, but general principles need to be tailored to the asylum context, drawing on legal standards designed to guard against outcomes that produce serious human rights consequences.
- Particular attention must be paid at national and regional level to how AI tools can support human rights-based decision-making in complex and politicized systems without exacerbating existing structural challenges. How we treat asylum seekers and refugees interacting with AI will be a test case for emerging domestic and regional legislation and governance of AI. Global standard-setting exercises for AI – including UN-based technical standards and high-level multinational initiatives – will also influence the direction of travel.

1. Introduction: The border of the future

The 'border of the future' is expected to be 'heavily dependent on digital systems, data analytics and automation-at-scale to both improve facilitation and mitigate risk'.[1] Work has already started on scoping out potential uses for artificial intelligence (AI) in immigration systems and asylum decision-making processes. For example, at the height of the COVID-19 pandemic, when many states restricted movement across borders, scenario-planning led by OECD member countries tackled the following near-future[2] proposition:

> What if, in 2035, many countries exploited advances in technology to select immigrants based on accurate and detailed assessments of their potential for successful integration and other desired characteristics? This could lead to better integration outcomes and great public support for migration. It could also give rise to debate about appropriate selection criteria, security, privacy and human rights concerns.[3]

Rapid shifts towards automation in various sectors, including at borders, raise questions about how to guarantee that international legal standards are carried through into the AI era.[4] This research paper offers a snapshot of the near-future outlook for AI in national systems that receive and process refugee protection claims, including when individuals seek asylum at borders, and associated concerns under international law. It explores emerging approaches to mitigating legal and ethical risks associated with AI in public decision-making, with the aim of supporting policymakers and civil society in thinking through effective safeguards in the asylum context and identifying gaps still to be addressed.

The appeal of AI and automated decision-making

There is no single definition of AI, but it can be usefully described as 'a set of computational technologies, that are inspired, but typically operate quite differently from, the way people use their nervous systems and bodies to sense, learn, reason, and take action'.[5]

---

[1] Michael Pezzullo, Australia's secretary of the Department of Home Affairs, quoted in Wroe, D. (2018), 'Top official's 'golden rule': in border protection, computer won't ever say no', *Sydney Morning Herald*, 15 July 2018, https://www.smh.com.au/politics/federal/top-official-s-golden-rule-in-border-protection-computer-won-t-ever-say-no-20180712-p4zr3i.html.
[2] An approach taken also by the International Committee of the Red Cross and the Oxford Future of Humanity Institute in prioritizing challenges associated with artificial intelligence.
[3] Organisation for Economic Co-operation and Development (2020), *Making Migration and Integration Policies Future Ready*, https://www.oecd.org/migration/mig/migration-strategic-foresight.pdf.
[4] Several key authors and commentators have addressed issues in this sector. See, for example, Beduschi, A. (2020), 'International migration management in the age of artificial intelligence', *Migration Studies*, 9(3), pp. 576–96; and Molnar, P. and Gill, L. (2018), *Bots at the gate: a human rights analysis of automated decision-making in Canada's immigration and refugee system*, Toronto: University of Toronto and The Citizen Lab.
[5] Stanford University (2016), *Artificial Intelligence and Life in 2030: One hundred year study on artificial intelligence, Report of the 2015 Study Panel*, California: Stanford University, p. 1. For a comprehensive overview of AI as a scientific discipline in policy development, and an explanation of key AI methods and their features, see Independent High-Level Expert Group in Artificial Intelligence (2019), *A*

The ability of AI to approximate human decision-making has created demand for 'automated' or 'algorithmic' processes that can support or act in the place of human decision-makers.[6] AI's appeal crosses many sectors in which decisions that may be informed by data are made, including within public administration.

Typically, AI-related technologies in the public sphere aim to improve the quality, accuracy, consistency, efficiency, effectiveness or timely delivery of functions. To date algorithms have been used more often to help decision-making that is high-volume and routine in nature, and more easily coded by expert systems.

Now, machine learning and other advanced and emerging techniques such as neural networks and natural-language processing are offering opportunities for AI to analyse vast quantities of data and identify patterns and correlations that can support strategic planning, inform investigations, and enable problem-solving in critical fields of government.[7] In other words, AI is becoming a feature of decision-making in situations that are inherently complex.

Why do asylum and refugee protection test AI?

The power of states to control their borders is tempered by their obligations under international law. Under international refugee[8] and human rights law,[9] states must not return individuals to countries where there are substantial grounds for believing they will face a real danger of persecution, torture or other serious human rights violations (this proscription under the principle of 'non-refoulement' includes pushbacks at borders). To prevent refoulement, states are expected to adopt a range of legal and practical interventions, such as establishing national systems (known as asylum systems) to assess claims to refugee status and other forms of international

definition of AI: Main capabilities and scientific disciplines, Brussels: European Commission, https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_definition_of_ai_18_december_1.pdf.

[6] The processes include those generating decisions that automatically trigger consequences (algorithm-determined); those generating decisions where human decision-making is required to take into account an algorithmic analysis, thus restricting the scope of the human decision (algorithm-driven); or those where decisions are informed by an algorithmic analysis (algorithm-based). See Daten Ethik Kommission [Data Ethics Commission] (2019), *Opinion of the Data Ethics Commission*, Berlin: Daten Ethik Kommission, https://assets.contentstack.io/v3/assets/blt3de4d56151f717f2/blt300ce23c9789e0f3/5e5cfe13fa08326331360f93/191023_DEK_Kurzfassung_en_bf.pdf.

[7] For an overview of different algorithmic techniques, including discussion of their possible uses and levels of interpretability, see UK Information Commissioner's Office and The Alan Turing Institute (2020), *Explaining decisions made with AI*, https://ico.org.uk/for-organisations/guide-to-data-protection/key-dp-themes/explaining-decisions-made-with-ai/.

[8] For an overview of relevant treaty provisions, see Nicholson, F. and Kumin. J. (2017), *A guide to international refugee protection and building state asylum systems: Handbook for Parliamentarians No. 27, 2017*, Inter-Parliamentary Union and UN High Commissioner for Refugees, https://www.unhcr.org/uk/publications/legal/3d4aba564/refugee-protection-guide-international-refugee-law-handbook-parliamentarians.html. Also see Article 33 of the Refugee Convention adopted 25 July 1951, noting limited exceptions to the principle of non-refoulement.

[9] For an overview of the scope of the principle under relevant treaty provisions, notably Article 3 of the UN Convention Against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment (Convention against Torture) (adopted 10 December 1984), as well as international customary law, see the brief publication: UN Office of the High Commissioner for Human Rights (2018), 'The Principle of non-refoulement under international human rights law', https://www.ohchr.org/Documents/Issues/Migration/GlobalCompactMigration/ThePrincipleNon-RefoulementUnderInternationalHumanRightsLaw.pdf.

protection in line with international legal standards, including the principle of non-discrimination.[10]

This means states need to design and administer government policies – including technological systems – in line with their legal obligations. This is a complex and politically sensitive task for many governments. The question of when states may limit entry at borders came into sharp relief during the COVID-19 pandemic, when a vast majority of countries imposed additional entry restrictions in order to 'health-proof' their borders.[11] These restrictions included the use of remote surveillance technologies, from temperature checks to location-tracking for quarantine, that in effect brought border security into people's homes.

The expansion of technology into health surveillance across borders follows decades of demand for more secure borders to combat terrorism and transnational crime, and for greater control over migration flows.

The expansion of technology into health surveillance across borders follows decades of demand for more secure borders to combat terrorism and transnational crime, and for greater control over migration flows. These factors, including the current pandemic, create 'moral panics'[12] that are often used to scapegoat migrants and refugees. The same factors are also contributing to an increase in measures to limit access to asylum, making it harder for asylum seekers to leave countries of risk and enter countries of safety, putting international law and the principle of non-refoulement under pressure. The constraints on access to asylum include extremely limited resettlement opportunities. As a result, migrants (including asylum seekers) are increasingly taking risky voyages to reach destination countries, including using people-smugglers.

This complex environment is a driver for technological innovation, and is renewing attention on how the principle of non-refoulement applies at borders.[13] The new and emerging technologies will operate in an environment where, if international legal standards are not rigorously

[10] UN Committee Against Torture (2017), 'General comment No. 4 (2017) on the implementation of article 3 of the Convention in the context of article 22', CAT/C/GC/4, p. 13.
[11] Kysel, I. M. (2020), ''Health-proofing' human mobility systems and new technologies of border control: An opportunity to advance the rights of all migrants?', Andrew & Renata Kaldor Centre for International Refugee Law, 27 August 2020, https://www.kaldorcentre.unsw.edu.au/publication/%E2%80%98health-proofing%E2%80%99-human-mobility-systems-and-new-technologies-border-control-opportunity.
[12] Tazreiter, C. and Metcalfe, S. (2021), 'New Vulnerabilities for Migrants and Refugees in State Responses to the Global Pandemic, COVID-19', *Social Sciences*, 10(9), 342, https://doi.org/10.3390/socsci10090342.
[13] This refers to activities that take place in the border region or transit zone, or even further afield, as part of 'border protection' measures as well as at the territorial border itself.

applied to AI tools and the ecosystems in which such tools are introduced, there may be real human consequences.[14]

Introducing AI systems into this field presents significant human rights-related challenges. At the same time, the existing legal protections against bias, unlawful decisions and refoulement are already under pressure. As the UN Special Rapporteur on contemporary forms of racism flagged when looking at the challenges of introducing new technology in this field: 'Executive and other branches of government retain expansive discretionary, unreviewable powers in the realm of border and immigration enforcement that are not subject to the substantive and procedural constraints typically guaranteed to citizens.'[15] These challenges can be exacerbated at scale when AI systems offer seemingly simple solutions to complex problems.[16]

This means that asylum and refugee protection will form one of the test cases for global and national governance of AI, and for whether human rights-compliant AI can be achieved. As the UN secretary-general has said: 'As refugees go, so goes the world.'[17]

2. The near future of AI and asylum

AI is well and truly on the radar for asylum and border authorities. At the request of the European Commission, in May 2020 the global advisory firm Deloitte identified a shortlist of AI capacities that could be operational within national asylum systems in the EU within five years.[18] However, recommendations from firms and the creation of pilot schemes trialling AI do not necessarily indicate that AI will become a key feature of asylum policy in the near future.[19] If legal and regulatory systems work as planned, not all of AI's potential uses will be implemented. In other words, a challenge for the AI era is to ensure that public authorities have not only the space and opportunity to investigate possible technical aids, but also the obligation to abandon initiatives that do not meet legal standards or the

[14] Molnar, P. (2019), 'Technology on the margins: AI and global migration management from a human rights perspective', *Cambridge International Law Journal*, 8(2), p. 306, https://doi.org/10.4337/cilj.2019.02.07; Beduschi (2020), 'International migration management in the age of artificial intelligence'.
[15] UN Special Rapporteur on contemporary forms of racism (2021), *Report of the Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance*, A/HRC/48/76, para. 3(b).
[16] Neff, G. (2020), 'AI @ Work: overcoming structural challenges to ensure successful implementation of AI in the workplace', Oxford Internet Institute, 13 August 2020, https://www.oii.ox.ac.uk/blog/ai-work-overcoming-structural-challenges-to-ensure-successful-implementation-of-ai-in-the-workplace/.
[17] Guterres, A. (2019), 'World must 'reboot' approach to refugees, first UN Global Forum hears', UN News, 17 December 2019, https://news.un.org/en/story/2019/12/1053671.
[18] Deloitte (2020), *Opportunities and challenges for the use of artificial intelligence in border control, migration and security*, pp. 24–26.
[19] Although algorithmic transparency is becoming more common, several studies – including one by the UK government's Committee on Standards in Public Life – have noted that 'public sector organisations are not sufficiently transparent about their use of AI and it is too difficult to find out where machine learning is currently being used in government'. Committee on Standards in Public Life (2020), *Artificial Intelligence and Public Standards*, London: Committee on Standards in Public Life, p. 6, https://www.gov.uk/government/news/artificial-intelligence-and-public-standards-committee-publishes-report.

measure of public trust (regardless of any financial pressures to continue with a particular solution).

What is clear is that as the range of available AI methods continues to grow, so does the range of possible interventions across the asylum decision-making cycle. Some current and expected near-future applications of AI in asylum systems are listed below, spanning two broad categories: decision-making support; and identity verification and risk analysis.

i. Decision-making support in asylum systems

In 1986, in an article aptly titled 'The British Nationality Act as a Logic Program', a research group tried to translate key legal standards in UK citizenship legislation into computer code. It was, they said, 'a rich domain for developing and testing artificial intelligence technology'[20] because the act contained 'vague' phrases such as 'being of good character' that were not defined in the legislation and would require factual and legal interpretation, making the standards difficult to put into code. When applied by decision-makers, these or similar vague phrases often carry unspoken values – such as what constitutes a 'well-founded fear' of being persecuted for refugee claims – or can be used in a way that discriminates based on race or other characteristics.

Today, AI can approximate some forms of human thinking and intelligence, but the technological capacity still does not exist to reliably code complex legal tests to determine a person's refugee status or need for protection against refoulement under international law. Assessments require decision-makers to have regard to the future possible risks to individuals refused entry or returned to their country of origin; such assessments also rely on complex and nuanced tests associated with confirming identity and credibility. Any effort solely reliant on AI to decide refugee status, or to reject claims for other forms of international protection based on future risk of human rights abuses, would be highly controversial.[21]

Table 1. Non-exhaustive examples of AI proposed for use in asylum systems to support decision-making

Purpose	Example of possible AI application
Identity verification	Germany's immigration authority, the Federal Office for Migration and Refugees, has piloted the use of digital tools, including facial and dialect recognition, to help verify personal identities within the

[20] Sergot, M. J. et al. (1986), 'The British Nationality Act as a Logic Program', *Communications of the ACM*, 29(5), pp. 370–86, http://www.doc.ic.ac.uk/~rak/papers/British%20Nationality%20Act.pdf, cited in Deeks, A. (2020), 'Coding the Law of Armed Conflict: First Steps' in Waxman, M. C. and Oakley, T. W. (eds) (2022, forthcoming),*The Future Law of Armed Conflict*, New York: Oxford University Press.
[21] Molnar and Gill (2018), *Bots at the gate*, p. 33. The authors explore various considerations in the Canadian context, including the opacity and discretionary nature of decision-making as a high-risk laboratory and an 'environment ripe for algorithmic discrimination'.

	asylum determination process when asylum seekers arrive in the country.[22] Other identification initiatives are noted below in Table 2.
Application triage and prioritization	The May 2020 Deloitte paper identifies the possibility of using AI-enabled micro-gesture and emotion analysis to support caseworkers responsible for identifying asylum seekers with particular vulnerabilities.[23] Vulnerability identification can be used to direct how an asylum seeker's claim is processed (such as in an expedited decision-making stream), and whether a person is referred to specialized medical, mental health or other services. It may also be taken into account in decisions relating to detention (e.g. assessing whether immigration detention is justified pending an asylum hearing or other process).
Assessing components of an application	Germany's 'Digitization Agenda' aims to achieve greater efficiency through end-to-end digitalization of workflows, with more ambitious goals for later stages to include AI to support processing and decision-making.[24] Other governments, as well as the above-mentioned Deloitte study, have investigated tools to analyse past judicial determinations so that decision-makers can better predict whether a decision to recognize or reject refugee status will be subject to appeal.[25] The Deloitte paper identifies the need for caution, but notes the potential for AI tools to generate and process the country-of-origin information used to help determine refugee status, via a system that can prepare country-of-origin-specific questions for decision-makers to rely on.

ii. Identity verification and risk analysis, including for border management

For many countries that receive and process millions of arrivals every year, efficient entry-processing systems are essential. Current AI initiatives are building on trends in the past decade towards the adoption of risk-based approaches to border management,[26] often supported by big data analytics.[27]

[22] Beduschi (2020), 'International migration management in the age of artificial intelligence'.
[23] Deloitte (2020), *Opportunities and challenges for the use of artificial intelligence in border control, migration and security*, pp. 24–26. The Deloitte report identifies the following possible methods: 'AI to perform real-time analysis of an applicant's facial movements, spoken language and body language to detect signals which can better inform decision-making by a human social worker/specialised expert (e.g. if the person should be granted special procedural guarantees)'. The report notes that some AI capacities such as emotional recognition are still 'nascent' or under development.
[24] Bundesamt für Migration und Flüchtlinge [German Federal Office for Migration and Refugees] (BAMF) (2018), *Digitisation Agenda 2020: Success stories and future digital projects at the Federal Office for Migration and Refugees (BAMF)*, Berlin: BAMF, p. 5. Stage III initiatives to deliver 'systematic support for decision making' include 'digital technologies such as data analysis or artificial intelligence used to support staff in a targeted manner when it comes to processing and decision making'.
[25] Ling, J. (2018), 'Federal government looks to AI in addressing issues with immigration system', *Globe and Mail*, 31 May 2018, https://www.theglobeandmail.com/politics/article-federal-government-looks-to-ai-in-addressing-issues-with-immigration/.
[26] Ajana, B. (2015) 'Augmented borders: Big Data and the ethics of immigration control', *Journal of Information, Communication and Ethics in Society*, 13(1), pp. 58–78.
[27] Ibid., p. 58.

Table 2. Non-exhaustive examples of AI proposed for use in identity verification and risk analysis

Purpose	Example of possible AI application
Identity verification	Facial recognition technology (FRT) is already used to verify identity in some jurisdictions.[28] FRT relies on facial biometrics. For example, e-passports use a photo of the passport-holder held on file that can be matched to a live image of the person presenting the passport (this is known as 'one-to-one matching'); the process is often accompanied by requirements to provide other biometrics such as fingerprints. FRT may also be used to match faces against profiles on watchlists of persons of interest or concern. This latter 'one-to-many' matching process typically uses machine learning to assess the probability that two images belong to the same person, and thus pick people out of crowds.
Risk assessment	Pilot projects in the US and EU have looked to AI (including advanced neural networks) for deception detection. A pilot EU 'Horizon 2020' project, iBorderCtrl, incorporated an avatar that interviewed travellers to analyse micro-gestures and non-verbal behaviour with the aim of identifying deception.[29] It was intended to operate alongside or within smart devices to support risk assessments. While the idea is that border officials would remain involved in final decisions to allow or deny entry, the accuracy and utility of these AI-based 'lie detector' tools in immigration processing are nonetheless highly contested.[30] The project has reportedly been scrapped,[31] and it is uncertain if it will be taken further. There is also an increasing reliance on complex algorithms to analyse large data sources against predetermined risk profiles. A UN Counter-Terrorism Centre (UNCCT) handbook explains profiling in this context as intended to 'assess whether an individual is a security risk or to help detect irregular migration'. In addition, the handbook states: 'Profiling means using information about a person to establish whether they are likely to pose a security or other risk. For example, factors such as travel from conflict zones may be used as part of a security risk assessment.

[28] Perry, M. (2019), 'iDecide: Digital Pathways to Decision', 2019 CPD Immigration Law Conference, Canberra, Law Council of Australia, https://www.fedcourt.gov.au/digital-law-library/judges-speeches/justice-perry/perry-j-20190321.
[29] iBorderCtrl (undated), 'Project Summary', https://www.iborderctrl.eu/ (not currently available).
[30] This is acknowledged by project partners. Ibid. 'Project Summary' webpage (not currently available) noted: '[I]t is important to emphasize that, while [the Automatic Deception Detection System] may reveal statistical likelihoods of deceptive behaviour, each case would require further checking by a human agent to determine if deception is individually present. This is also a legal requirement under the EU General Data Protection Regulation and law enforcement directive (prohibition on automatic decision-making). ADDS is based on previously developed technology, in particular the so-called Silent Talker. The iBorderCtrl project has adopted this technology but is also well aware of the scientific controversy around its efficacy.' See also Varghese, S. (2018), 'The science behind the EU's creepy new border tech is totally flawed', *Wired*, 16 November 2018, https://www.wired.co.uk/article/border-control-technology-biometrics; and Katwala, A. (2019), 'The race to create a perfect lie detector – and the dangers of succeeding', *Guardian*, 5 September 2019, https://www.theguardian.com/technology/2019/sep/05/the-race-to-create-a-perfect-lie-detector-and-the-dangers-of-succeeding.
[31] Heikkilä, M. (2021), 'Europe's Artificial Intelligence Blindspot: Race', *Politico*, 16 March 2021, https://www.politico.eu/article/europe-artificial-intelligence-blindspot-race-algorithmic-harm/.

	It may be done either automatically through the collection of data remotely or in person, for example, through questioning at the border or consideration of the information provided on landing cards.'[32]
Detention assessments	Algorithmic systems are used for public safety and flight risk assessments for immigration detention decisions in the US.[33] Such systems are listed in the Deloitte study as offering the opportunity 'to predict risk of an applicant absconding during review of application and the return process (e.g. using variables such as country of origin, previous application history, age)'.[34]
Surveillance, front-end and back-end applications	Research for Frontex, the European Border and Coast Guard Agency, identified a range of applications including both 'front-end' capabilities, which end-users would directly utilize (e.g. security gates and surveillance systems), and 'back-end' capabilities, which would have an enabling impact on border security functions (e.g. automated machine learning).[35]

3. Challenges in meeting international legal standards

The most common ethical and legal challenges associated with the use of AI in asylum and related border and immigration systems involve issues of opacity and unpredictability, the potential for bias and unlawful discrimination, and how such factors affect the ability of individuals to obtain a remedy in the event of erroneous or unfair decisions. There are also questions of legal liability and accountability: that is, who is responsible when things go wrong? In domains such as asylum, any of the above issues in decision-making processes can directly impinge on rights, with irreversible effects.

To properly capture these factors and move beyond general concerns surrounding AI, risks can usefully be broken down into what Yeung identifies as 'outcomes-based' and 'process-based' categories (although these are not neat distinctions, as procedural standards help ensure outcomes

[32] UN Counter-Terrorism Centre (2018), *Handbook on human rights and screening in border security and management*, UN Office of Counter-Terrorism, https://www.un.org/sites/www.un.org.counterterrorism/files/1806953-en-ctitf-handbookhrscreeningatborders-for-web2.pdf.
[33] Koulish, R. (2016), 'Immigration Detention in the Risk Classification Assessment Era', *Connecticut Public Interest Law Journal*, 16(1).
[34] Deloitte (2020), *Opportunities and challenges for the use of artificial intelligence in border control, migration and security*, p. 24.
[35] Frontex (2021), *Artificial intelligence-based capabilities for the European Border and Coast Guard: Final Report*, https://frontex.europa.eu/assets/Publications/Research/Frontex_AI_Research_Study_2020_final_report.pdf.

strike a legitimate balance between different rights and interests).[36] The exact risk for a given application will still depend on the nature of the tool used, what it has been developed for, the details of training provided, how the tool is used, and who uses it.

Outcomes-based risks

Potential outcomes-based risks that will often need to be considered prior to implementing AI tools in asylum systems include the following:

1. **Malicious use.** There is an absolute need to guard against malicious use of an AI tool resulting in human rights violations. This could include, for example, the deliberate use of facial recognition technology (FRT) for surveillance where this amounts to unlawful interference in privacy,[37] or the deployment of FRT and AI capacities that prevent persons facing a real fear of persecution from leaving their own country at official borders to seek international protection.[38]
2. **Physical safety risks.** These may arise if 'smart' border technology incorporating AI applications reduces access to asylum or produces changed patterns of travel or behaviour by asylum seekers or people-smugglers.[39]
3. **Potential for refoulement.** Strict safeguards are needed to ensure that any AI tool intended to support the delivery of policy goals, including security imperatives at or around borders, does not result in outcomes that contravene the principle of non-refoulement. For example, international law prohibits automatic pushbacks at borders without consideration of the travellers' personal circumstances, including the opportunity for individuals to inform authorities of their need for asylum and seek an assessment of their claim. Over the coming years, tools that progressively automate border control will have to be assessed against their potential impact on the ability of states to continue to meet these obligations.[40]

[36] Yeung, K. (2019), 'Why Worry about Decision-Making by Machine?' in Yeung, K. and Hodge, M. (2019), *Algorithmic Regulation*, Oxford: Oxford University Press.
[37] For an overview of the standards for lawful surveillance under international human rights law, see, for example, UN High Commissioner for Human Rights (2018), *The right to privacy in the digital age*, Report of the United Nations High Commissioner for Human Rights, A/HRC/39/29.
[38] UN General Assembly (1966), International Covenant on Civil and Political Rights, Article 12(2). Although practices limiting the ability of women to exit a territory without the permission of a guardian preceded the introduction of new digital technologies, the role of apps and the potential future role of other technology are subjects of ongoing human rights concern. See, for example, Human Rights Watch (2019), 'Saudi Arabia's Absher App: Controlling Women's Travel While Offering Government Services', 6 May 2019, https://www.hrw.org/news/2019/05/06/saudi-arabias-absher-app-controlling-womens-travel-while-offering-government.
[39] Chambers, S. N., Boyce, G. A., Launius, S. and Dinsmore, A. (2019), 'Mortality, Surveillance and the Tertiary "Funnel Effect" on the U.S.-Mexico Border: A Geospatial Modeling of the Geography of Deterrence', *Journal of Borderlands Studies*, 36(3), pp. 443–68, cited in UN Special Rapporteur on contemporary forms of racism (2021), *Report of the Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance*.
[40] UN Special Rapporteur on contemporary forms of racism (2021), *Report of the Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance*, para. 11. The publication cites reports regarding differential treatment, due to errors in facial recognition technology, which 'frequently perpetuates negative stereotypes and may even entail prohibited discrimination that could lead to refoulement'.

4. **Accuracy problems.** An AI tool – such as a natural-language processing tool[41] or FRT – may produce an inaccurate result that ultimately contributes to inaccurate decision outcomes for individuals, with potentially serious human rights and legal consequences.[42]

5. **Discriminatory outcomes.** Higher rates of inaccurate results could occur 'at the margins' for individuals or groups subject to structural or systemic inequality.[43] For example, false positives[44] are often higher for non-Caucasian faces in FRT algorithms developed in Europe and the US. This presents 'privacy and civil rights and civil liberties concerns such as when matches result in additional questioning, surveillance, errors in benefit adjudication, or loss of liberty'.[45]

6. **Perpetuation of systemic discrimination and marginalization**. Profiling and predictive tools rely on past patterns of observed behaviour among groups of people to make decisions about individuals – including, in some cases, about their anticipated behaviour in the future. Such tools are used, for example, in immigration detention decisions in the US to determine whether an irregular immigrant who applies for asylum is likely to flee or cause harm in the community, and thus whether detention is needed or not. The UN High Commissioner for Human Rights has cautioned that 'predictive tools carry an inherent risk of perpetuating or even enhancing discrimination' because the past data used to make predictions will often 'reflect racial and ethnic bias'[46] or carry harmful assumptions and stereotypes.

---

[41] Ibid., p. 11: 'Such algorithms have an important place in migration management as they can be used for dialect recognition, streamlining asylum determination processes […] [But] [g]aps in the data about the dialects of ethnic minorities used to train the algorithms could reinforce existing patterns of discrimination vis-à-vis these ethnic minorities.'

[42] Beduschi (2020), 'International migration management in the age of artificial intelligence', p. 7: 'AI algorithms may accidentally misidentify a migrant as a terrorist or miscalculate the risk of ill-treatment upon deportation to their country of origin. Blind over-reliance on AI technologies could lead to serious breaches of human rights if in these scenarios, migrants were deprived of liberty due to misidentification, or if they were subjected to torture or inhuman treatment upon deportation.'

[43] The principles of equality and non-discrimination are codified in international human rights law, such as articles 2(1) and 26 of the International Covenant on Civil and Political Rights requiring that rights set out in that covenant are recognized without distinction of any kind, such as race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status. Note that *differentiation* in treatment may be permitted where this serves a legitimate objective under the human rights treaty in question and the criteria applied are reasonable and objective.

[44] For a simple explanation of false positives and false negatives using the example of an email spam filter that classifies emails, the UK Information Commissioner's Office (ICO) offers: '[…] a false positive or 'type I' error: these are cases that the AI system incorrectly labels as positive (eg emails classified as spam, when they are genuine); or a false negative or 'type II' error: these are cases that the AI system incorrectly labels as negative when they are actually positive (eg emails classified as genuine, when they are actually spam)'. UK ICO (2020), *Guidance on the AI auditing framework: Draft guidance for consultation*, p. 48, https://ico.org.uk/media/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf.

[45] The US Department of Commerce's National Institute of Standards and Technology (NIST) notably tested algorithms against four available US government-held datasets: domestic 'mugshots', application photographs for immigration benefits, visa photographs and border crossing photographs. The NIST notes that the first three sets 'have good compliance with image capture standards'. However, it also noted that the last (border crossing photographs) did not, given 'constraints on capture duration and environment'. Together, these datasets allowed the NIST to process 18.27 million images of 8.49 million people through 189 mostly commercial algorithms from 99 developers. National Institute of Standards and Technology (2020), 'Evidence of Dr. Charles Romaine before the Committee on Homeland Security, United States House of Representatives', 6 February 2020, https://www.nist.gov/speech-testimony/facial-recognition-technology-frt-0.

[46] Office of the UN High Commissioner for Human Rights (2021), *The right to privacy in the digital age: Report of the United Nations High Commissioner for Human Rights*, A/HRC/48/31, para. 24.

Box 1. Untangling systemic bias and racism – human or machine?

A cautionary tale can be found in the existing use of algorithms to carry out risk assessments for immigration detention. US researchers found that human decision-makers were more likely to override a computer-generated recommendation not to detain (a release recommendation) and to issue a decision to impose immigration detention than they were to override a computer-generated recommendation to detain (a detention recommendation) and recommend release pending immigration hearings.

This tendency for human decision-makers to be more risk-averse than the tool's recommendation will often act as a feedback loop informing subsequent updates to the system, as tended to occur in the above-mentioned example.[47]

This means that policy decisions about the level of deference to be afforded to human decision-making as a guide for AI learning development, and technical specifications as to how a given AI system learns, will also influence the extent to which AI can be relied upon in human rights-sensitive uses. Tracking and untangling responsibility for any bias or trend in decision-making (is it the human or the machine determining the outcome?) will become an even more central component of litigation and public accountability.

Process-based risks

As above, potential process-based issues and problems that may need consideration prior to implementing AI tools in asylum systems include the following:

1. **Bias, unfairness or unlawful discrimination.** There is a significant risk of bias, unfairness or unlawful discrimination corrupting design and implementation processes. Internal rules or logic may undermine an AI system's ability to exclude 'irrelevant considerations' – among these are attributes such as race, protected under domestic and international laws on non-discrimination. Related risks exist around processes that control for proxy indicators of race or protected attributes where consideration of these factors is not reasonably justified.[48] As the granting of refugee protection will also typically be based on identity factors (e.g. persecution on the grounds of race), the challenge for system design will be to distinguish between relevant and irrelevant protected factors within any AI tool used in asylum-related decision-making.

[47] Koulish, R. and Evans, K. (2020), 'Injustice and the disappearance of discretionary detention under Trump: Detaining Low Risk Immigrants without Bond', Interdisciplinary Laboratory of Computational Social Science Working Paper, 22 May 2020.
[48] See above on legitimate distinctions which do not amount to discrimination.

2. **Failure to consider individual circumstances in assessments.** AI-enabled profiling capacity and other AI tools may prevent an individualized assessment from being carried out in respect of a claim of asylum or appeal against deportation when this is required by domestic or international law.

Box 2. Individualized decision-making and the challenge of profiling

Particular care is needed in designing or introducing AI capacities that rely on profiling (i.e. group-based analysis). For example, profiling could be used to predict the future behaviour of individuals, thereby informing decisions about whether immigration detention is required where domestic national legislation may provide for detention in some cases.

Over the years, reliance on individualized decision-making in asylum systems has constituted a primary safeguard against mistakes and discrimination, including against preferential treatment based on race. For example, EU directives require applications for international protection to be assessed 'individually, objectively and impartially'.[49] The European Convention on Human Rights (ECHR) also prohibits collective expulsion of aliens from a territory without consideration of their personal circumstances.[50] Even legal tests that are 'forward-looking' require an assessment on the basis of evidence relating to the individual.[51]

Individualized decision-making is also a key safeguard against arbitrary immigration detention (such detention being contrary to the right to liberty under international human rights law).[52] Immigration detention should be used only as an exceptional measure of last resort 'justified as reasonable, necessary and proportionate in the light of the circumstances specific to the individual case'.[53]

Existing tools that use group attributes have often fallen short of this standard. For example, the lawfulness of psychological assessments in asylum cases is often contested, as seen in European case law and in the views of the Advocate-General:

> If I understand correctly, the hidden conflicts or emotions that such an analysis is supposed to uncover would, in the eyes of the psychologists carrying it out, either confirm or call into question the applicant's stated sexual orientation. It would seem to me, though, that such type of analysis inevitably involves the use of stereotyped notions as to the behaviour of homosexuals. […] That is, a type of analysis that the Court has already found problematic in [previous case law] *A and Others*, insofar as it does not permit

[49] Directive 2013/32/EU, Article 10(3)(a).
[50] Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights, as amended), Article 4 of Protocol No. 4. See European Court of Human Rights (2020), *Guide on Article 4 of Protocol No. 4 to the European Convention on Human Rights*, https://www.echr.coe.int/Documents/Guide_Art_4_Protocol_4_ENG.pdf.
[51] This would also include exceptions to the principle of non-refoulement in international refugee law under Article 33 of the Refugee Convention, including on the basis of prospective threat to national security. See further Goodwin-Gill, G. S. and McAdam, J. (2007), *The Refugee in International Law 3rd edition*, Oxford: Oxford University Press, p. 241.
[52] A v Australia, Communication No 560/1993 UN Doc. CCPR/C/59/D/560/1993 (3 April 1997) para. 9.4.
[53] UNHCR (2012), *Detention Guidelines*, UNHCR, p. 14, https://www.unhcr.org/505b10ee9.html.

full account to be taken of the individual situation and personal circumstances of the applicant.[54]

So long as AI-enabled capacities rely on group-based or past historic cases, their exclusive use in government decision-making will often fall short of international legal standards where individualized assessments are expected.[55] This is a concern that has been explored – and affirmed – in recent US case law on the use of AI in the criminal justice sphere.[56]

3. **The challenge of maintaining high procedural fairness standards.** National asylum systems that assess eligibility for international protection are required to meet standards of 'fairness, efficiency, adaptability and integrity'[57] as well as domestic administrative and regional asylum law standards.[58] However, there are already severe shortcomings in many domestic national systems.[59] AI technologies generate a range of challenges for meeting procedural fairness standards. For example, in many types of case – including decisions about refugee status – the decision-maker needs to give written reasons for any decision to deny refugee status.[60] This is intended to give the applicant the opportunity to understand why their claim was rejected and to seek an effective remedy before a court or a tribunal; this requires particular attention to be paid to the 'explainability' of AI used in the decision-making process.[61]

   Molnar and Gill also point to a range of challenges that can compromise AI tools. For example, tools designed to help decision-makers identify factors determining success (i.e. which case should be successful, which decision is likely to be overturned on appeal compared to other similar cases) may sound useful. But the deployment of such tools may not be

[54] Opinion of Advocate General Wahl, Case C-473/16 F v Bevándorlási és Állampolgársági Hivatal. [2017], Opinion of Advocate General Wahl (2017), para. 37, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62016CC0473.
[55] McGregor, L., Murray, D. and Ng, V. (2019), 'International human rights law as a framework for algorithmic accountability, *International & Comparative Law Quarterly*, Volume 68, Issue 2, pp. 309–43.
[56] State of Wisconsin v Eric L. Loomis [2016] WI 68 (13 July 2016), https://www.wicourts.gov/sc/opinion/DisplayDocument.pdf?content=pdf&seqNo=171690. In the criminal justice sphere, the Supreme Court of the State of Wisconsin in the US allowed algorithmic risk assessments to be put before the judges setting sentencing conditions so long as they were not the 'determinative' factor. Sole use of the AI-enabled analysis, the court concluded, would 'raise due process challenges regarding whether a defendant received an individualized sentence' because 'COMPAS risk assessment scores are based on group data, they are able to identify groups of high-risk offenders—not a particular high-risk individual'. Importantly, the court confirmed that risk analytics should not be admitted as a consideration for whether a custodial sentence should be applied, or for how long. Such decisions were to be assessed based on the severity of the crime committed rather than future risk, e.g. of reoffending. The latter might be tempting but could not be justified to 'sentence offenders to more time than they morally deserve'.
[57] Executive Committee of the High Commissioner's Programme (1977), 'Determination of Refugee Status No. 8 (XXVIII)', UNGA Document No. 12A (A/32/12/Add.1); UNHCR UK, 'Refugee Status Determination', https://www.unhcr.org/uk/refugee-status-determination.html.
[58] Oswald, M. (2018), 'Algorithm-assisted decision-making in the public sector: framing the issues using administrative law rules governing discretionary power', *Philosophical Transactions of the Royal Society A*, Volume 376, Issue 2128, https://royalsocietypublishing.org/doi/10.1098/rsta.2017.0359.
[59] UN Special Rapporteur on contemporary forms of racism (2021), *Report of the Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance*, para. 2.
[60] See, for example, Directive 2013/32/EU, Article 11(2).
[61] Ibid., Article 46.

appropriate if, in past cases, success typically depended on an asylum seeker's access to legal advice and representation, which is not always guaranteed in many jurisdictions; in such instances, this could mean that an AI tool learns what success looks like based on a skewed existing system.[62]

4. **Difficulty ensuring that use of personal data meets privacy and data protection standards**.[63] The right to privacy is an end in itself, as well as a key 'entry point' for considering the impact of digitalized decision-making on the enjoyment of human rights.[64] One of the key selling points for many AI technologies is their ability to identify patterns or anomalies in large datasets drawn from multiple sources. But this also creates concerns about interference with personal privacy because, as Calo puts it, AI can make visible the 'intimate from the available'[65] or may be able to generate inferences about a person's past or future behaviour.[66]

5. **Unlawful interference with privacy.** There will often be legitimate reasons for authorities to request and verify personal information in visa processes, at national borders and during asylum proceedings.[67] Nevertheless, data collection and analysis may unlawfully interfere with privacy if the data or collection processes are not necessary for, or proportionate to, the achievement of a legitimate aim. Determining the lawfulness of such operations requires consideration both of a tool's effectiveness and of alternative, less invasive methods to verify identity and the credibility of claims. Establishing the effectiveness of new or experimental methods will present particular difficulty.[68] Moreover, AI tools often process large amounts of data from different sources. This can easily raise questions of intrusive overreach against privacy standards, especially as more databases become interlinked (in configurations known as interoperable databases).[69]

[62] Molnar and Gill (2018), *Bots at the gate*, p. 55.
[63] Article 17, International Covenant on Civil and Political Rights.
[64] See, notably, reliance on Article 8 (the right to respect for private and family life) under the European Convention on Human Rights in examining an AI tool, including for discriminatory impact, in the Hague District Court, Nerderlands Juristen Comite voor de Menssenrechten et al v The State of the Netherlands, ECLI: NL: RBDHA: 2020: 1878 (the SyRI decision).
[65] Calo, R. (2017), 'Artificial Intelligence Policy: A Primer and Roadmap', *University of California, Davis Law Review*, Volume 51, pages 399–404.
[66] Wachter, S. and Mittelstadt, B. (2019), 'A right to reasonable inferences: Re-thinking Data Protection Law in the Age of Big Data and AI', 2019 *Columbia Business Law Review*, Volume 2019, Issue 2, pages 494–620.
[67] UNHCR (2017), 'Preliminary Legal Observations on the Seizure and Search of Electronic Devices of Asylum-Seekers', 4 August 2017, https://www.refworld.org/publisher,UNHCR,POSITION,,59a5231b4,0.html.
[68] Barrett, L. F. et al. (2019), 'Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements', *Psychological Science in the Public Interest*, Volume 20, Issue 1, pp. 1–68, https://journals.sagepub.com/doi/pdf/10.1177/1529100619832930.
[69] Bunyan, T. (undated), 'EU: Interoperability of Justice and Home Affairs databases: a "point of no return"', Statewatch, https://www.statewatch.org/observatories/eu-interoperability-of-justice-and-home-affairs-databases-a-point-of-no-return/.

6. **Expanded data sharing.** Any AI-associated increase in data sharing, for example between state and non-state actors, will require a well-defined framework with sufficient safeguards to protect against the unlawful or arbitrary use of data, including in cases pertaining to asylum seekers and their families. A commonly cited concern is the potential for personal data to be inadvertently passed to an individual asylum seeker's country of origin, alerting authorities to their claim of persecution.[70]

Box 3. Challenges for system-based effects and evidentiary standards

An underexplored challenge in the development of AI is to understand and anticipate how the technology will work across a system – in particular, how its use may benefit one part of a larger, complex decision-making system while causing harm in another part of the system. A further element of the challenge is working out how to prevent such harms.

Many AI tools offer probabilities only. A facial recognition tool, for example, does not provide 100 per cent certainty that one face matches another, as it relies on an assessment of how similar one image is to another. Similarly, AI-enabled deception detection methods at border control points may not be designed to achieve absolute accuracy.[71] Rather, they may be designed to help authorities with a limited number of human border control officers 'reduce the haystack' of persons arriving at a border so that deception detection can focus on nominated 'high-risk' travellers.[72] When such tools are used across large groups, the 'scalability of AI solutions can dramatically increase negative effects of seemingly small error rates'.[73]

Fundamental judgment on the appropriateness of using AI to support security risk assessments is beyond the scope of this paper. Nonetheless, it is possible to note that where tools only offer a probability of something, they should not be assumed to be accurate. And, where a tool maintains a tolerance for false results, it is critical to understand how that tool may influence other, later steps in a decision-making process.

For example, if an AI deception detection machine flags an asylum seeker as suspected of having presented false documents or made misrepresentations to officials at a border entry point, this can trigger the use of accelerated procedures for assessing the asylum claim in some jurisdictions.[74] These accelerated

[70] Huszti-Orbán, K. and Ní Aoláin, F. (2020), *Use of Biometric Data to Identify Terrorists: Best Practice or Risky Business?*, Minneapolis: Human Rights Centre, University of Minnesota, p. 9, https://www.ohchr.org/Documents/Issues/Terrorism/Use-Biometric-Data-Report.pdf.
[71] Katwala (2019), 'The race to create a perfect lie detector – and the dangers of succeeding'.
[72] Ajana (2015), 'Augmented borders: Big Data and the ethics of immigration control'.
[73] OHCHR (2021), *The right to privacy in the digital age: Report of the United Nations High Commissioner for Human Rights*, A/HRC/48/31, 13 September 2021, para. 18.
[74] European Union, Directive 2013/32/EU of the European Parliament and of the Council on common procedures for granting and withdrawing international protection (recast), Article 31(8).

measures, while common, are often criticized for failing to preserve procedural fairness standards and for being misapplied in ways that can easily discriminate.[75]

Any AI tools that contribute, whether intentionally or not, to decisions to funnel some applicants into an accelerated procedure must therefore be treated with extreme caution.

4. Emerging solutions

Setting 'red lines'

AI governance is a global work in progress. From a human rights perspective, the current rate of progress in the development of AI governance brings its own risks of harm, as the pace of innovation and use is outstripping that of regulatory change.

The UN High Commissioner for Human Rights has called for a moratorium on AI systems that pose serious risks to human rights until adequate safeguards, including legislative protections, are in place. A moratorium could include tools that have not been sufficiently tested and checked for discriminatory outputs,[76] and 'black box' applications that affect administrative or judicial review of decisions affecting the legal rights of individuals and the right to an effective remedy.[77] Tools that interfere with other rights, such as the right to privacy, in ways that are unlawful or not clearly necessary and proportionate to the achievement of public policy goals would also be non-rights-compliant and should not be permitted. The commissioner has also called for a permanent ban on AI applications that cannot be used in compliance with international human rights law.[78]

In domestic legislation, so far only the EU has put forward draft law that can prohibit the production, sale and use of human rights-non-compliant AI.[79] Unveiled in April 2021 and still under negotiation, the European Commission's proposal would see some uses for AI prohibited because of their inherent likelihood of breaching fundamental rights.

[75] European Council on Refugees and Exiles (2017), *Accelerated, prioritised and fast-track asylum procedures: Legal frameworks and practice in Europe*, https://www.ecre.org/wp-content/uploads/2017/05/AIDA-Brief_AcceleratedProcedures.pdf.
[76] R (on the application of Edward Bridges) v the Chief Constable of South Wales Police and the Secretary of State for the Home Department [2020] EWCA Civ 1058.
[77] McGregor, Murray and Ng (2019), 'International human rights law as a framework for algorithmic accountability', pp. 309–43.
[78] OHCHR (2021), *The right to privacy in the digital age, Report of the United Nations High Commissioner for Human Rights*, A/HRC/48/31 (13 September 2021), https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=27469&LangID=E.
[79] Other jurisdictions developing AI-specific legislation – including Brazil and the UK – do not: Coalizão Diretos (2021), 'Na Rede Brasil não está pronto para regular inteligência artificial' [Brazil is not ready to regulate artificial intelligence], 7 December 2021, https://direitosnarede.org.br/2021/12/07/brasil-nao-esta-pronto-para-regular-inteligencia-artificial/; HM Government (2021), *National AI Strategy*, Command Paper 525, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1020402/National_AI_Strategy_-_PDF_version.pdf.

At the same time, the current list of prohibited items in the draft EU law is extremely narrow: essentially limited to tools that have been proven to cause physical harm or that apply subliminal techniques likely to cause psychological harm. The draft EU AI Act contemplates regulating, but not prohibiting, AI for emotion recognition and lie detection at borders – even though this is one of the most controversial use cases trialled to date, one for which the scientific basis is also questionable. Such an application of AI therefore ought to be a candidate for a moratorium or ban, given its expected interference with freedom of thought and privacy.[80]

The draft EU AI Act contemplates regulating, but not prohibiting, AI for emotion recognition and lie detection at borders – even though this is one of the most controversial use cases.

The draft AI Act also fails to include overarching protections such as a prohibition on tools that cannot meet the full scope of human rights standards under European law (including principles of non-discrimination), or to make clear that any interferences with rights must meet legal tests of necessity and proportionality for narrowly defined purposes. Domestic legislation being introduced in other jurisdictions also falls short of imposing moratoriums or bans on selected uses.[81]

Sector-specific risk-based regulation

Instead of moratoriums, many national and regional governments including the EU are prescribing higher levels of scrutiny for designated 'high-risk' AI uses. Known as risk-based regulation, several legislative proposals – including from Canada and the EU – recognize explicitly that AI in asylum systems carries risk.[82] But these proposals contain limited guidance about how to manage risk in a way that meets human rights obligations when designing AI for this sector.

The draft EU AI Act[83] goes furthest by listing specific high-risk domains in an annex to the draft legislation. Systems classed as high-risk AI will be expected to meet certain pre- and post-application conditions that are broadly aligned with European privacy regulations.[84] Several potential

[80] EDPB-EDPS (2021), *Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*, https://edpb.europa.eu/system/files/2021-06/edpb-edps_joint_opinion_ai_regulation_en.pdf.
[81] The UK is expected only to include some limits on FRT. Brazil draft legislation is principles-based only.
[82] Government of Canada (undated), 'Algorithmic Impact Assessment tool', https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html.
[83] European Commission (2021), 'Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts', COM/20201/206.
[84] Ibid., proposed articles 9–15. These include criteria based on the EU's General Data Protection Regulation around risk management, documentation, transparency, human oversight, accuracy, robustness and cybersecurity.

applications of AI in 'migration, asylum and border control management' are explicitly recognized as high-risk: notably, tools to assess the security risk of individuals; systems 'intended to assist [...] the examination of applications for asylum, visa and residence permits'; and more controversial systems that were not subject to a ban, such as polygraphs and tools to detect emotional states.[85]

However, in July 2022 the Czech presidency of the Council of the EU released amendments to the draft that would narrow the scope of the high-risk category to only AI uses in high-risk sectors that have 'a significant bearing on a decision or immediate effect' on individuals. Given that in asylum systems even ancillary uses for AI such as triaging can have an impact on a final outcome for an individual's claim, the sector as a whole should be treated with caution. If the changed language is adopted, 'significant bearing' and 'immediate effect' must be interpreted broadly, to account for the fact that AI tools can have a significant impact on rights even when there is a human in the loop or non-automated methods are used as well.[86]

Moreover, while risk-based management is common in environmental and social risk assessments, its suitability for setting standards for automated government decision-making remains open to question[87] unless government agencies (or companies performing public functions) are required to demonstrate risk mitigation to recognized legal standards, not some other level of risk tolerance.

Box 4. Outstanding questions for risk-based approaches to AI in asylum contexts

Canada's Algorithmic Impact Assessment helps government departments to select which oversight and review processes are needed for a particular AI tool based on four tiers of risk. At the lowest level – Level I – an AI system or automation may not present significant risks and may therefore require only minimal approvals. At the highest level – Level IV – an AI tool operates in the context of government decisions that are deemed to be 'high stakes', and that 'will often lead to impacts [on individuals] that are irreversible and are perpetual'.[88]

The Canadian process alerts departments to the range of risks that need to be measured and tackled. But terms such as 'high stakes' are not inherently in line with legal and human rights standards. Legislative direction may be required to

[85] Ibid., Annex 3.
[86] Council of the European Union (2022), *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts - Second Presidency compromise text*, Interinstitutional File: 2021/0106(COD), https://artificialintelligenceact.eu/wp-content/uploads/2022/07/AIA-CZ-1st-Proposal-15-July.pdf; McGregor, L., Murray, D. and Ng, V. (2019), 'International human rights law as a framework for algorithmic accountability, *International & Comparative Law Quarterly*, Volume 68, Issue 2, pp. 309–43
[87] Human Rights Watch (2021), 'How the EU's Flawed Artificial Intelligence Regulation Endangers the Social Safety Net: Questions and Answers', 10 November 2021, https://www.hrw.org/news/2021/11/10/how-eus-flawed-artificial-intelligence-regulation-endangers-social-safety-net; Access Now (2021), 'The EU should regulate AI on the basis of rights, not risks', 17 February 2021, https://www.accessnow.org/eu-regulation-ai-risk-based-approach/.
[88] Government of Canada (2019), 'Directive on Automated Decision-Making', 5 February 2019.

ensure that the questions posed to government departments, and the responses to them, are consistent with existing legal and policy standards. As the Canadian Bar Association submitted to the national immigration, refugee and citizenship authorities: 'Would automated decision-making on cases involving vulnerable persons always be classified as Level IV decisions?'[89]

Meaningful human control

To override and avert some of the risks associated with fully automated decision-making, many policymakers emphasize the importance of human control over AI systems. This is evident, for instance, in the European Commission's draft AI Act.[90] There has also been an identifiable shift in AI policy thinking towards augmenting rather than replacing human action and expertise, given questions about whether AI tools can apply the same logic as humans, and questions about trust and accountability (reflecting the sense that a human or institutional subject should remain accountable and liable for outcomes).

This position is replicated in public statements from some immigration authorities. Australian officials, for example, have publicly affirmed that even where AI technologies are used to inform a decision, a human decision-maker will make the final determination.[91] UK officials have also given evidence that data-based risk profiling is not used in immigration to make *solely* automated decisions,[92] and that procedures retain a 'human in the loop'. Products for border security functions are often developed on the proviso that they will provide advice and assistance only to a final human decision-maker.[93]

But the conditions for meaningful and effective human control are still being worked out, including through the courts. One case with relevance for the asylum sphere occurred in 2017, when the European Court of Justice examined a proposed EU–Canada agreement to authorize the cross-border transfer of airline-collected passenger data (known as Passenger Name Records or PNRs) for security pre-screening purposes. The bilateral agreement anticipated that the data would be assessed using automated methods (including algorithms, but not necessarily advanced AI). Given that the data collected could then be used by a human decision-maker to make

[89] The Canadian Bar Association (2019), 'Re: Artificial Intelligence and Machine Learning in Immigration Law', Letter from Sedai, M. to Hussen, M., 11 July 2019, https://www.cba.org/CMSPages/GetFile.aspx?guid=c54903f5-cd8a-4d3a-96a3-ce0c33623845.
[90] See Fjeld, J. et al. (2020), *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI*, Cambridge, MA: Berkman Klein Center for Internet & Society, p. 53, with particular emphasis on human review of fully automated decisions identified in 69 per cent of documents in the dataset.
[91] Wroe (2018), 'Top official's 'golden rule': in border protection, computer won't ever say no', quoting Australian secretary of the Department of Home Affairs, Michael Pezzullo: 'If you are denied a visa, you will be denied by a human officer. They might be prompted by an AI, they might be assisted by AI, but it's a human that will deny your visa. We call that the 'golden rule' [...] No robot or no artificial intelligence system should ever take away someone's right, privilege or entitlement in a way that can't ultimately be linked back to an accountable human decision-maker.'
[92] R (Open Rights Group & the3million) v Secretary of State for the Home Department [2019] EWHC 2562 (Admin) [20].
[93] See, for example, project documentation for iBorderCtrl.

binding decisions about an individual's authority to enter Canada, the court provided a non-exhaustive list of recommendations to ensure the data transfer agreement met EU data protection standards, including in the following areas:

- **Non-interference with privacy:** Pre-established models and criteria should be specific and reliable, making it possible to arrive at results targeting individuals who may be under a 'reasonable suspicion' of participation in terrorist offences or serious transnational crime. The results should be 'non-discriminatory' to prevent indiscriminate interference with the right to privacy.
- **Human control:** Given the potential for error, any 'positive result' must be subject to an individual re-examination by non-automated means before the relevant action adversely affects the air passenger in question.
- **Reliability:** The reliability of automated models must be subject to review under the EU–Canada agreement.[94]

While principles-based requirements for information exchange are welcome, there is increasing consensus that effective human control will be difficult to achieve where an AI tool contains complex algorithms that operate as a 'black box' system. In such cases, a tool may not be fit for purpose (and so should be subject to a ban or moratorium) until solutions such as 'explainable AI' are well developed.[95] The UK Information Commissioner's Office has recommended that organizations looking to use AI tools should select tools that can already be reviewed easily for accuracy or where the logic of the rules used can be explained.

Human control is also not a perfect protection against harm. To act as a safeguard against mistakes with significant consequences, effective human control needs to be just that – effective. Determining factors in this respect include the decision-maker's expertise and capacity to consider, review and make decisions that are appropriately informed by, but independent of, an AI analysis.[96] This in turn depends on what non-AI-derived information is available, the scope of the human decision-maker's legal and actual authority to reject AI-generated results, and the human decision-maker's level of professional knowledge.[97] Many national asylum systems already suffer from under-resourcing and limited training for decision-makers,[98]

[94] *Official Journal of the European Union* (2017), 'Opinion of the Court (Grand Chamber) of 26 July 2017 – European Parliament (Opinion 1/15)', 2017/C 309/3. Similar concerns are now being litigated in relation to the equivalent EU directive that governs passenger data sharing within the EU, and the EU is in the process of renegotiating the agreement with Canada.
[95] Artificial Intelligence Committee, UK House of Lords (2017), *AI in the UK: Ready, Willing and Able?*, Report of Session 2017-9, 16 April 2017, para. 105: 'We believe it is not acceptable to deploy any artificial intelligence system which could have a substantial impact on an individual's life, unless it can generate a full and satisfactory explanation for the decisions it will take. In cases such as deep neural networks, where it is not yet possible to generate thorough explanations for the decisions that are made, this may mean delaying their deployment for particular uses until alternative solutions are found.'
[96] State of Wisconsin v Eric L. Loomis (2016).
[97] Loomis (2016), para. 70–74.
[98] UNHCR (2018), 'Asylum Capacity Support Group: Note for Discussion', https://www.unhcr.org/5cc1aba44.pdf.

leaving rejected asylum claims vulnerable to being overturned on appeal.[99] Other common characteristics of asylum systems, such as the potential for political pressure on decision-makers, limit the suitability of relying on a 'human in the loop' to act as a control mechanism for automated functions.

Legislative safeguards

Courts are generally sympathetic to public authorities' reasons for wanting to introduce AI systems, but have ruled against governments in several recent cases because of lack of legal safeguards and legislative oversight. Courts are placing the burden back on authorities and providers of AI to justify the need and work through legislative oversight processes, as opposed to individuals having to challenge AI systems through the courts.[100]

For example, UK police were trialling FRT in public spaces for identification of persons on watchlists. In deciding that FRT did not meet standards of legality, the Court of Appeal found flaws with the process to authorize the use of FRT and identified a lack of clear, authorized criteria for deciding how individuals would be selected and placed on a watchlist, which would then be fed into the FRT system.[101] Similarly, the existence of border watchlists and stop-lists populated mostly by individuals from minority communities has also raised concerns about unlawful discrimination.[102]

UK courts have also recently affirmed that the cost of creating and maintaining a technological system is not a sufficient rationale for failing to correct policy, legislation and technological unfairness.[103]

Responsible innovation

Canadian and Australian immigration authorities have both explored the possibility of automating so-called 'neutral' or 'positive' decisions in immigration visa application triaging and identification matching.[104] The logic for this is that AI tools would not be used to make a final (or near-final) decision that could negatively affect a person's legal rights or obligations, thereby potentially generating a reason to seek review of the decision under domestic legal frameworks. Used in a non-discriminatory way, AI assistance in the triage and prioritization of cases in decision-making streams could

[99] See, in relation to the UK, a civil society analysis of historic reports and recommendations: Freedom from Torture (2019), *Lessons not Learned: The failures of asylum decision-making in the UK*, https://www.freedomfromtorture.org/sites/default/files/2019-09/FFT_LessonsNotLearned_Report_A4_FINAL_LOWRES_1.pdf.
[100] Molnar, P. (2020), *Technological Testing Grounds: Migration Management Experiments and Reflections from the Ground Up*, EDRi (European Digital Rights), https://edri.org/wp-content/uploads/2020/11/Technological-Testing-Grounds.pdf.
[101] R (on the application of Edward Bridges) v the Chief Constable of South Wales Police and the Secretary of State for the Home Department [2020] EWCA Civ 1058 at paras 199 and 90–91 respectively.
[102] UN Special Rapporteur on contemporary forms of racism (2021), *Report of the Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance*, para. 22.
[103] Secretary of State for Work and Pensions v Johnson and others, [2020] EWCA Civ 777.
[104] Parliamentary Joint Committee on Intelligence and Security (2019), *Advisory report on the Identity-matching Services Bill 2019 and the Australian Passports Amendment (Identity-matching Services) Bill 2019*, Parliament of the Commonwealth of Australia, https://parlinfo.aph.gov.au/parlInfo/download/committees/reportjnt/024343/toc_pdf/AdvisoryreportontheIdentity-matchingServicesBill2019andtheAustralianPassportsAmendment(Identity-matchingServices)Bill2019.pdf;fileType=application%2Fpdf.

speed up the processing of asylum claims – to the benefit of both refugees and governments.

Caution may be needed, though, as intermediary steps (such as triage and prioritization) in asylum processes can themselves produce negative consequences and be subject to appeal.[105] The UK recently removed algorithms from use in immigration triage after legal actions seeking their disclosure for bias testing were commenced, although the authorities denied the tools were biased.[106] Separately, a Dutch court found that AI used to identify risk factors for welfare fraud that would then be investigated by a human could – unless safeguards were put in place, including greater transparency and testing for discrimination – still arbitrarily interfere with the right to privacy under European human rights law.[107]

The UK recently removed algorithms from use in immigration triage after legal actions seeking their disclosure for bias testing were commenced, although the authorities denied the tools were biased.

Multi-stakeholder development?

In debates on responsible and trustworthy AI, significant emphasis is placed on the need for multi-stakeholder engagement to create AI that is more likely to be beneficial, lawful and ethical, and that does not build in risks from the outset. Such debates consider the use of multi-disciplinary teams in the design of AI processes, and the potential for co-creation or consultation with stakeholders to avoid the introduction of human rights blind spots *ex ante*.

There have been comparable efforts in the justice sector, including from judges, to identify uses for AI that would receive broad support within the sector (including from civil society), and to identify AI tools that present 'red flags', in order to steer policymakers on the types of tools that should or should not receive investment.[108]

There are several other potentially significant benefits to the co-creation of AI systems in the asylum sector. Firstly, such systems, if well designed, could support dignified application procedures in what is currently an

[105] UNHCR (2010), *Improving asylum procedures: comparative analysis and recommendations for law and practice: Detailed Research on Key Asylum Procedural Directive Provisions,* https://www.unhcr.org/4c7b71039.pdf. The report notes the number of countries where a decision to prioritize or apply an accelerated process is challengeable.
[106] McDonald, H. (2020), 'Visa applications: Home Office refuses to reveal 'high risk' countries', *Guardian,* 1 January 2020, https://www.theguardian.com/uk-news/2020/jan/01/visa-applications-home-office-refuses-to-reveal-high-risk-countries.
[107] Nerderlands Juristen Comite voor de Menssenrechten et al v The State of the Netherlands, ECLI: NL: RBDHA: 2020: 1878 (the SyRI decision).
[108] European Commission for the Efficiency of Justice (CEPEJ), *European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment*, adopted at the 31st plenary meeting of the CEPEJ (Strasbourg, 3–4 December 2018), https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c.

adversarial, contested and undignified space.[109] Members of displaced communities trying to enter a country for safety are often desperate and have suffered trauma. Overly intrusive searches for data in phones, social media and other sources to create risk profiles also cause asylum seekers to eschew technology that can and should help them to remain connected to friends and family.[110]

Secondly, private firms are more alert than ever to their human rights duties and impacts on vulnerable communities. Effective business human rights due diligence should be informed by consultation with potentially affected groups.[111] The UN Guiding Principles on Business and Human Rights urge businesses to pay special attention to human rights impacts on individuals from groups or populations that may be at heightened risk of vulnerability or marginalization; this requires a tailored understanding of the rights of those groups and an appreciation of the risks generated by AI in the specific contexts in which automated processes will be used.[112]

But the highly changeable nature of asylum policy – along with trends towards more restrictive borders, as seen at the height of the COVID-19 pandemic – is likely to stymie collaboration, creating real risks for community groups and technology firms that might wish to partner with governments. Policy shifts towards more restrictive immigration regimes may well create legal and reputational questions for private sector partners that design, license and operate AI-based systems,[113] given corporate responsibility to respect international human rights law standards[114] as well as domestic privacy and non-discrimination legal regimes.

For example, since 2013, the US Immigration and Customs Enforcement (ICE) agency has used a computerized risk classification assessment to help determine whether to detain or release a person pending an immigration (including asylum) hearing. As an existing tool embedded within an established system, ICE's computerized assessment has presented researchers with the challenge of improving its operational effectiveness. Proposals have included the potential integration of predictive analytics that can better account for 'equity factors', including the rate at which asylum is ultimately granted.[115] In theory this could assist officers in reducing the number of people detained. However, according to researchers during several US administrations, the variables and weightings have been manipulated to adapt the tool's outputs to meet the immigration policy agenda prevailing at any given time – including the former Trump administration's 'no release' policy on the mandatory detention of illegal immigrants.[116] The result has effectively been to remove the tool's option to

[109] Molnar (2020), *Technological Testing Grounds*.
[110] Meaker, M. (2018), 'Europe is using smartphone data as a weapon to deport refugees', *Wired*, 2 July 2018.
[111] OHCHR (2011), *Guiding Principles on Business and Human Rights*, A/HRC/17/31.
[112] Principle 18.
[113] Molnar and Gill (2018), *Bots at the gate*, p. 59, citing employee and public pushback at large technology firms supplying (non-AI) products to US Immigration and Customs Enforcement systems.
[114] OHCHR (2011), *Guiding Principles on Business and Human Rights*.
[115] Koulish (2016), 'Immigration Detention in the Risk Classification Assessment Era'.
[116] Executive Order 13768 (2017), revoked by executive orders issued under President Joe Biden.

recommend against detention.[117] Arguably, this has undermined the goals of a risk-based approach, which in theory should allow individuals who are considered low-risk to remain in the community pending immigration proceedings.

The dilemma for business is clear: although tools must be able to adapt to changing official policy, this adaptability poses significant challenges. Should firms accept that if government policy becomes increasingly anti-immigration and runs contrary to international legal obligations, technology tools should follow suit?

Should firms accept that if government policy becomes increasingly anti-immigration and runs contrary to international legal obligations, technology tools should follow suit?

Human rights compliance for businesses performing state functions

Governments are increasingly outsourcing refugee protection, border management and AI at the same time. This creates a complex 'government–business' nexus in which private entities have significant involvement in, and sometimes control over, the design and implementation of policy but do not necessarily have the same level of accountability in terms of respecting and protecting rights.

The UN Office of the High Commissioner for Human Rights (OHCHR) has put it bluntly and in practical terms. If states are going to rely on the private sector to deliver public goods or services, they have to be able to oversee such processes and demand accuracy and transparency around human rights risks. If not satisfied that the risks can be mitigated, states should not use private contractors to deliver public goods or services.

A non-exhaustive list of tools that can increase confidence in private sector delivery of public services via AI systems includes the following:

1. Mandatory human rights impact assessments

Smart regulatory mixes for AI already include risk and impact assessments. But as indicated above, these tools will not be able to look at all risks to individuals in a way that meets international legal obligations. To address the issue, in Europe the Committee of Ministers of the Council of Europe has advocated the use of compulsory human rights impact assessments (HRIAs) for all public sector AI systems, in addition to any data protection, social,

[117] Robertson. A. (2020), 'ICE rigged its algorithms to keep migrants in jail, claims lawsuit', The Verge, 3 March 2020, https://www.theverge.com/2020/3/3/21163013/ice-new-york-risk-assessment-algorithm-rigged-lawsuit-nyclu-jose-velesaca.

economic or other impact assessments required under existing law.[118] Standalone HRIAs or assessments embedded in other tools are also relevant for businesses – which have their own duties under multilateral frameworks and domestic and regional legal systems[119] – as well as for international organizations.[120]

To truly mitigate AI-related risks in the context of asylum processes, assessments need to be broad enough to address system-based effects, and – as the Committee of Ministers has put it – include an evaluation of the 'possible transformations that these systems may have on existing social, institutional or governance structures'.[121] Timing is also critical. Assessments should occur 'regularly and consultatively' throughout the design and deployment processes, notably 'prior to public procurement, during development, at regular milestones, and throughout their context-specific deployment in order to identify the risks of rights-adverse outcomes'.[122]

2. Third-party audits and ongoing independent review functions

Complementing HRIAs, the proposed EU AI Act requires developers to prove compliance with certain standards such as on accuracy. But it allows developers and users to self-certify and self-monitor such compliance. This approach takes into account the likelihood that a vast number of AI systems will become a part of everyday life, making it difficult to demand and regulate independent third-party review or oversight of all systems.

However, it will remain crucial to have third-party audits and independent oversight in some domains, including asylum. Advocates of safeguards have argued that the opaque, discretionary and often discriminatory nature of decision-making in asylum and border control, along with the growing role of for-profit private entities in government, demands independent oversight.[123]

In addition, independent reviews and supervisory functions can alleviate the burden on individuals to challenge AI-related assessments, particularly when the people involved may already be constrained by limited access to domestic justice mechanisms, a lack of resources, and language barriers. The New Zealand Algorithm Charter for government use of AI recommends the use of peer review for algorithms and encourages departments to 'act

[118] Council of Europe Committee of Ministers (2020), 'Recommendation of the Committee of Ministers to member States on the human rights impacts of algorithmic systems'.
[119] Proposed mandatory EU-wide supply chain due diligence, European Parliament (2021), 'European Parliament resolution of 10 March 2021 with recommendations to the Commission on corporate due diligence and corporate accountability (2020/2129(INL))', www.europarl.europa.eu/doceo/document/TA-9-2021-0073_EN.html.
[120] UN Special Rapporteur on contemporary forms of racism (2021), *Report of the Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance*, A/HRC/48/76.
[121] Council of Europe Committee of Ministers (2020), 'Recommendation of the Committee of Ministers to member States on the human rights impacts of algorithmic systems', para. 5.2.
[122] Ibid.
[123] Molnar, P. (2022), 'Territorial and Digital Borders and Migrant Vulnerability Under a Pandemic Crisis', Triandafyllidou, A. (ed.) (2022), *Migration and Pandemics: Spaces of Solidarity and Spaces of Exception*, IMISCOE Research Series, https://link.springer.com/content/pdf/10.1007/978-3-030-81210-2.pdf.

on' the results of those reviews.[124] This tends in the right direction, but the effectiveness of this voluntary charter has not yet been tested. Nor is it an alternative to ensuring mandatory access to legal remedies for those affected by wrongful decisions.

3. Minimum viable fairness and accuracy levels for products

The next stages in AI policy development will need to define the level of acceptable risk against human rights standards. This is not easy.

Computer science research is actively looking to improve the accuracy, verifiability and reliability of AI tools after they leave the training environment. There are even hopes that AI can help to eliminate profiling based on generalized assumptions relating to race, ethnicity or other factors.[125] The argument is that if a tool can search in a discriminatory way for patterns that indicate risk, it should also be able to look for patterns that *reveal* discrimination; however, some doubt that 'fair learning AI' is really feasible.[126]

There are hopes that AI can help to eliminate profiling based on generalized assumptions relating to race, ethnicity or other factors.

For public sector applications being developed or introduced now and in the near future, something more will be needed. Courts have said clearly that they expect public bodies introducing new technologies to 'satisfy themselves that everything reasonable which [can] be done [has] been done' to prevent unlawful bias or other flaws.[127] This includes products obtained from private vendors or developers.

Meeting this requirement is also challenging for public sector entities. As the UK's Centre for Data Ethics and Innovation highlighted in a public enquiry:

> Public servants are likely to face significant trade-offs between different kinds of fairness and between fairness and accuracy. There is currently limited guidance and a lack of consensus about how to make these choices or even how to have constructive and open conversations about them. These choices are likely to be highly context specific.[128]

[124] New Zealand Government (2020), *Algorithm Charter for Aotearoa New Zealand*, https://data.govt.nz/assets/data-ethics/algorithm/Algorithm-Charter-2020_Final-English-1.pdf.
[125] See, for example, Kaye, D. (2018), *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, UN Doc A/73/348, 29 August 2018, para. 6.
[126] Some commentators looking at criminal justice tools say that it is unrealistic to try to address legal, social and political issues related to fairness through computation. See, for example, Green, B. (2018), 'Fair Risk Assessments: A Precarious Approach for Criminal Justice Reform', paper presented at 5th Workshop on Fairness, Accountability, and Transparency in Machine Learning (FAT/ML 2018), Stockholm, Sweden, https://www.fatml.org/media/documents/fair_risk_assessments_criminal_justice.pdf.
[127] See R v The Chief Constable of South Wales Police and others [2020] EWCA Civ 1058 [199].
[128] Centre for Data Ethics and Innovation (2019), 'Response to the Committee on Standards in Public Life's AI & Public Standards Review', 31 July 2019, p. 8.

The challenges for asylum systems are likely to mirror those in other domains. Are trade-offs between accuracy and transparency lawful and justifiable? Can a level of accuracy and assurance against discrimination be maintained over time, and how would this be monitored and evaluated? How would problems be rectified?

The 'right' level of transparency

There is currently a strong focus on defining transparency standards for public sector AI,[129] with some encouraging indicators in evidence but a lot of questions still to be addressed.

The draft EU AI Act proposes a public register for high-risk AI uses, with notice to be provided to the people and entities affected. This replicates calls by advocates for a public register of AI systems used in asylum cases.[130] In the UK, the government is expected to accept the recommendation of the national Commission on Race and Ethnic Disparities mandating transparency for all public sector organizations 'applying algorithms that have an impact on significant decisions affecting individuals'.[131] In New Zealand, the government department responsible for asylum and immigration has signed up to a New Zealand Algorithm Charter (July 2020) requiring departments to maintain transparency by 'clearly explaining how decisions [that affect individuals] are *informed* by algorithms', with recommendations on how to achieve this.[132] But in all cases, the scope of the obligation (e.g. is a decision 'informed by' AI if the AI performs a sorting function?) has yet to be defined and tested.

For the proposed EU AI Act to meet EU data protection standards, transparency requirements should be sufficient to allow for independent review and should apply to both final decisions and 'intermediary' processes (such as vulnerability assessments that rely on profiling).[133] Recent EU jurisprudence has highlighted the critical importance to final outcomes of transparency about AI used within a broader system, and its critical importance to the risk of human rights harm. A Dutch court was not given access to a system used to help identify welfare recipients who should be investigated for welfare fraud. Without some level of access to 'independently verifiable information' about how the system worked, the court found a violation of the right to privacy under EU law, because it was impossible to assess whether the interference was necessary and proportionate. Consequently, the court found that there was at least a possibility of discriminatory interference in the private lives of welfare recipients. The fact that the tool did not itself produce a final decision did

[129] Angwin, J., Larson, J., Mattu, S. and Kirchner, L. (2016), 'Machine Bias', ProPublica, 23 May 2016, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.
[130] Molnar (2020), *Technological Testing Grounds*.
[131] HM Government (2021), *National AI Strategy*.
[132] New Zealand Government (2020), *Algorithm Charter for Aotearoa New Zealand*.
[133] The EU's General Data Protection Regulation (GDPR) includes specific transparency requirements, the effectiveness of which is still being debated, wherein automated systems are used 'solely' to make legal or similarly significant decisions about individuals. See Bygrave, L. (2019), 'Minding the Machine v2.0: The EU General Data Protection Regulation and Automated Decision-making' in Yeung, K. and Hodge, M. (2019), *Algorithmic Regulation*, Oxford: Oxford University Press.

not alter the court's assessment, particularly as the tool was being trialled in economically deprived neighbourhoods.[134]

Big questions that remain include what exceptions to evolving transparency standards should be permitted. For example, when can authorities use AI without notifying individuals or providing access to information about how AI assessments were made (and on the basis of what data)?

Even without the introduction of AI, the permissible scope of limitations on data protection[135] – and on secrecy provisions around the disclosure of evidence – is still being litigated in a number of jurisdictions, including in relation to data and evidence in immigration deportation proceedings within Europe.[136] Introducing automated decision-making into systems veiled with secrecy is fraught, and there is a hard-fought debate about the right to information and ability to challenge decisions.

Any carve-outs or exceptions to AI transparency introduced in legislation should be strictly limited. They must be narrow, align with legal standards, and not ultimately undermine avenues for independent and judicial review, so that those affected can still assert their rights and seek remedy where necessary.

Striking a balance between rights and interests is likely to become more, not less, complex with the introduction of AI technologies.

Striking a balance between rights and interests is likely to become more, not less, complex with the introduction of AI technologies. Particular factors to consider will include the use of predictive analytical tools based on profiling, the reliance of automated tools on increasingly large datasets controlled by private sector actors, and the increasing presence of AI across large-scale, complex IT and decision-making systems.[137]

[134] Ibid., paras 6.6–6.7, 6.47 and 6.91.
[135] Europe Street News (2020), 'EU parliament warns about data transfer risks due to UK immigration rules', 16 February 2020, https://europestreet.news/eu-parliament-warns-of-data-transfer-risks-due-to-uk-immigration-rules/.
[136] UN Special Rapporteur on the promotion and protection of human rights and fundamental freedoms while countering terrorism (2019), 'Submission by the United Nations Special Rapporteur on the promotion and protection of human rights and fundamental freedoms while countering terrorism in *Muhammad and Muhammad v Romania* (Application No. 80982/12) before the European Court of Human Rights', 22 July 2019, https://www.ohchr.org/EN/Issues/Terrorism/Pages/AmicusBriefsExpertTestimony.aspx.
[137] Annex IX of the European Commission's proposed AI Act includes carve-outs from the overarching AI regulatory obligations – including transparency requirements – for large-scale IT systems in European migration and border control where a system is already placed on the market or put into service and subsequent change is not deemed significant.

5. Global efforts on AI

Because technological adoption can spread quickly and widely across borders, tackling the ethical and legal concerns attached to AI will not be achieved by domestic efforts only. Although most multilateral AI governance initiatives are in their infancy,[138] standard-setting for AI in the asylum sector is poised to develop significantly in three notable areas: multilateral data sharing frameworks; high-level initiatives; and development assistance in migration contexts.

Multilateral frameworks for data sharing

Multilateral frameworks that set the terms for data exchange between states are beginning to include guidance and minimum standards for automation and AI. The frameworks include those on information exchange in immigration, which also has implications for refugee protection. Pioneering efforts at standard-setting can be helpful, but they can place a significant burden on less resourced countries that seek to adopt AI early, potentially without the necessary human and legal safeguards against harm.

For example, the International Civil Aviation Organization (ICAO) recently revised standards for the collection and analysis of Passenger Name Records (PNRs). PNRs are used by airlines for a variety of commercial purposes and to facilitate implementation of UN Security Council resolutions relating to the prevention of terrorism and the use of risk-based security controls for airline passengers.[139]

The revised ICAO standards take on board a European Court of Justice 2017 opinion and the EU data protection regime (cited above in the section 'meaningful human control').[140] This is a positive step, recommending that states 'base the automated processing of PNR data on objective, precise and reliable criteria that effectively indicate the existence of a risk, without leading to unlawful differentiation', and that they discourage 'decisions that produce significant adverse actions affecting the legal interests of individuals based solely on the automated processing of PNR data'.[141]

These 'legal interests' should logically include the ability to seek asylum and protection against refoulement in both transit and destination countries. However, given the expanding number of situations in which cross-border data exchange will rely on automated processing, protecting freedom of movement and the ability to leave a place of risk in an automated risk-assessment era requires further attention. Data sharing without safeguards can also place asylum seekers and their families at harm

[138] For example, the UN Secretary-General's advisory group on AI (UN Secretary General Roadmap for Digital Cooperation, June 2020) and the Global Partnership on AI initiated by Canada and France.
[139] Including UN Security Council Resolution 2396 (2017), operative paragraph 12.
[140] European Council (2019), 'Council Decision (EU) 2019/2107 of 28 November 2019 on the position to be taken on behalf of the European Union within the Council of the International Civil Aviation Organization as regards the revision of Chapter 9 of Annex 9 (Facilitation) to the Convention on International Civil Aviation in respect of standards and recommended practices on passenger name record data', https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32019D2107.
[141] Liu, F. (2020), letter from Secretary-General ICAO to member states regarding adoption of Amendment 28 to Annex 9, https://acsa.cocesna.org/wp-content/uploads/2020/07/071-ENG.pdf.

if the information shared reveals that they have left their country of origin and sought asylum against persecution. Other very practical questions concern how to ensure all states are equipped to assess risks objectively, prevent bias, identify political or other interests in data and machine learning, and provide meaningful opportunities for human appeal, oversight and interventions. This seems all the more necessary given that emerging technologies will be promoted by large, well-resourced early adopters of AI through bilateral and multilateral frameworks.

High-level initiatives on AI

Numerous initiatives have begun the search for common ground on ethical principles for AI. Many of these aim to bring ethical principles together into normative frameworks, and to reinforce commitment to existing human rights law and standards so as to provide a legal foundation for ethical considerations.[142] Some initiatives are expected to focus on technology, such as FRT, used at borders. A perceived need to counter the rise in China's capacity to export surveillance technologies, including FRT, was an impetus for the US joining the Global Partnership on AI (GPAI) with like-minded states in June 2020.[143]

Multilateral efforts may help to promote minimum standards for AI where serious human rights concerns are associated with particular technologies and actors. However, commentators rightly fear that high-level principles may fall short of providing enforceable rights and safeguards for non-citizens (such as asylum seekers) when translated into domestic frameworks, given already high levels of public tolerance for new technologies at borders.[144]

Development assistance frameworks for migration management

There are long-standing debates about whether development aid can – or should – be linked to specific goals of donor countries, including the management and reduction of refugee and migrant flows.[145] Meanwhile, technological assistance for immigration infrastructure, border management and refugee systems is already common.[146]

Donor countries and international organizations offering technological assistance are required to exercise due diligence to ensure assistance and

[142] See, for example, UNESCO (2020), 'Major progress in UNESCO's development of a global normative instrument on the ethics of AI', 17 September 2020, https://en.unesco.org/news/major-progress-unescos-development-global-normative-instrument-ethics-ai.
[143] Chafkin, M. (2020), 'U.S. Will Join G-7 AI Pact, Citing Threat From China', Bloomberg, 28 May 2020, https://www.bloomberg.com/news/articles/2020-05-28/g-7-ai-group-adds-u-s-citing-threat-from-china?srnd=technology-vp.
[144] Molnar (2019), 'Technology on the margins'.
[145] Latek, M. (2019), *Interlinks Between Migration and Development*, EPRS Briefing, No. PE 630.351, European Parliamentary Research Service.
[146] See recommendations to the International Organization for Migration set out in UN Special Rapporteur on contemporary forms of racism (2021), *Report of the Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance*, A/HRC/48/76, para. 59.

cooperation do not result in human rights violations abroad. Appropriate due diligence includes the transparent and effective use of HRIAs.[147]

As bilateral or multilateral assistance begins to incorporate new technological functions, attention will need to be paid to how these applications operate at an individual and system-wide level. Among other aspects, due consideration will need to cover any impacts on global movement of people – especially when technical assistance is coupled with agreements on the sharing of personal data.

Box 5. AI as a 'moonshot' vision in prospects for reform of the global refugee system

With so many countries in the Global South hosting disproportionately large refugee populations, there are regular demands for new ways to promote the sharing of responsibility for people displaced by conflict and persecution. This includes appeals for more refugee humanitarian visa places to third countries and quicker, more efficient processing of refugee and humanitarian visas.

Technology cannot resolve fundamental structural challenges for the global refugee system. Still, the creative potential bound up with AI continues to inspire what could be characterized as 'moonshot' visions for possible reform, including of the global resettlement system.

James Hathaway, an eminent professor of international refugee law, has offered an ambitious reform roadmap with AI at its core, based on pilot projects in the US and Switzerland that use AI to match refugees with potential resettlement locations (e.g. towns or cities) based on a range of factors including economic or job opportunities.[148] Under Hathaway's proposals, AI would allow the preferences of destination states and refugees to 'be factored into a resettlement assignment system, relying on sophisticated algorithms to generate speedy matches'.[149]

Hathaway acknowledges that his vision is unlikely to be realized in the current political climate.[150] If such tools are politically compromised, overly restrict personal autonomy or demonstrate bias in outputs, they may further compromise global cooperation, responsibility sharing and ultimately refugee integration. This is a particular risk if refugees are placed in under-resourced and marginalized locations.[151] The sentiment, however, that a better way is needed to find alignment between the needs of displaced communities and the concerns of others, and that AI may be able to help with this, is worth exploring.

[147] Ferstman, C. (2020), 'Human Rights Due Diligence Policies Applied to Extraterritorial Cooperation to Prevent "Irregular" Migration: European Union and United Kingdom Support to Libya', *German Law Journal*, 21(3), p. 459, https://doi.org/10.1017/glj.2020.29.
[148] Immigration Policy Lab (2019), 'Implementing the Algorithm', https://immigrationlab.org/2019/04/11/implementing-the-algorithm/.
[149] Hathaway, J. C. (2018), 'The Global Cop-Out on Refugees', *International Journal of Refugee Law*, 30(4), p. 597, https://academic.oup.com/ijrl/article/30/4/591/5310192.
[150] UNHCR (2018), *Fair and Fast: UNHCR Discussion Paper on Accelerated and Simplified Procedures in the European Union*, 25 July 2018, https://www.refworld.org/docid/5b589eef4.html.
[151] Molnar and Gill (2018), *Bots at the gate*, p. 39.

6. Conclusion

Achieving human rights-compliant technology at the intersection of global politics and international law in asylum settings will be a test case for global and national governance of AI. Right now the risks – including those associated with the shortcomings of new technology, and of human and legal systems – outweigh the available protections. The introduction of AI should therefore be treated with extreme caution.

In the future, the viability of any given AI intervention in the asylum sector will depend significantly both on technological capacity and on deliberate human effort to build trust and accountability into new systems. The latter will only be achieved through collaborative design that prioritizes human rights-compliant AI and outcomes; pays careful attention to immediate and long-term human rights risks; provides independent oversight to review decisions and offer effective remedies; and keeps technology agile, adaptable to a changeable policy context, and removable if found to contain flaws.

Much will also turn on how asylum seekers and refugees are perceived and treated by the societies and legal systems in which they are seeking protection. As many AI techniques, including machine learning, are developed through learning from past decisions and behaviours, the test bed for future AI applications is how decisions are made regarding asylum and refugee status now.

The 'moonshot' vision for reform of the refugee regime nevertheless offers some hope that AI can inspire solutions to complex and highly politicized challenges. It suggests that a mantra for AI in this field could be 'optimism in innovation, legal protection of rights first before application'.

About the author

Madeleine Forster is a Chatham House Queen Elizabeth II Academy associate, and previously the Richard and Susan Hayden Academy Fellow 2019–20, hosted by the International Law Programme.

She brings expertise in applied international human rights law across complex emerging challenges and political, security and humanitarian environments, including refugee contexts. She is currently policy lead for inclusive climate action at the C40 Cities Climate Leadership Group. Prior to joining Chatham House, Madeleine provided specialist international legal services to United Nations humanitarian operations in the Middle East.

Acknowledgments

The author wishes to thank the Queen Elizabeth II Academy for Leadership in International Affairs, and Richard and Susan Hayden, for the unique opportunity afforded by the fellowship that made this paper possible.

Thanks also to the anonymous peer reviewers for their time, valuable comments and reflections. Sincere thanks to Ruma Mandal, Chanu Peiris and Harriet Moynihan within the Chatham House International Law Programme for their warm welcome, expertise and support. Thanks to Jake Statham for publication support and to Marjorie Buchser and the Chatham House Digital Society Initiative. Views expressed are the author's only.

Cover image: An aerial photo shows a boat carrying migrants stranded in the Strait of Gibraltar before being rescued by the Spanish Guardia Civil and the Salvamento Maritimo sea search and rescue agency, September 2018.
Photo credit: Copyright © Marcos Moreno/Contributor/AFP/Getty Images

ISBN 978 1 78413 532 4
DOI 10.55317/9781784135324

Cite this paper: Forster, M. (2022), *Refugee protection in the artificial intelligence era: A test case for rights*, Research Paper, London: Royal Institute of International Affairs, https://doi.org/10.55317/9781784135324.

This publication is printed on FSC-certified paper.
designbysoapbox.com

Independent thinking since 1920

The Royal Institute of International Affairs
Chatham House
10 St James's Square, London SW1Y 4LE
T +44 (0)20 7957 5700
contact@chathamhouse.org | chathamhouse.org

Charity Registration Number: 208223